Durham University

Durham Research Online

Deposited in DRO:

29 July 2016

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

da Silva-Junior, A.H.M. and Einbeck, J. and Craig, P.S. (2016) 'Gradient test for generalised linear models with random effects.', in Proceedings of the 31st International Workshop on Statistical Modelling. July 4-8, 2016, Rennes, France. , pp. 213-218.

Further information on publisher's website:

Publisher's copyright statement:

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full DRO policy for further details.

Durham University Library, Stockton Road, Durham DH1 3LY, United Kingdom
Tel : +44 (0)191 334 3042 | Fax : +44 (0)191 334 2971
https://dro.dur.ac.uk

Gradient test for generalised linear models with random effects

Antonio Hermes Marques da Silva Junior[12], Jochen Einbeck[1], Peter S. Craig[1]

[1] Durham University, Durham, UK
[2] Universidade Federal do Rio Grande do Norte, Natal, Brazil

E-mail for correspondence: a.hermes.marques-da-silva-jun@durham.ac.uk

Abstract: This work develops the gradient test for parameter selection in generalised linear models with random effects. Asymptotically, the test statistic has a χ^2 distribution and the statistic has a compelling feature: it does not require computation of the Fisher information matrix. Performance of the test is verified through Monte Carlo simulations of size and power, and also compared to the likelihood ratio, Wald and Rao tests. The gradient test provides the best results overall when compared to the traditional tests, especially for smaller sample sizes.

Keywords: Generalised linear models; random effects; asymptotic test.

1 Generalised linear models with random effects

Consider a generalised linear model with random effects (GLMwRE) for a data set containing n independent observations of a response variable, denoted $\mathbf{y} = (y_1, \ldots, y_n)^\top$, which by definition has a distribution in the exponential family, and corresponding observations on p explanatory variables, denoted $\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{ip})^\top$ for $i = 1, \ldots, n$. The linear predictor for the i-th observation is $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} + z_i$ where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is the vector of regression parameters and z_i is an unobserved random effect. The relationship between y_i and η_i is given by the conditional mean $\mu_i = \mathrm{E}[y_i|z_i]$ and the monotonic and differentiable *link function*, $g(\,\cdot\,)$ such that $\mu_i = g^{-1}(\eta_i)$. The z_i can be considered as sampled from $\mathcal{N}(0, \sigma^2)$, where $\sigma > 0$. An alternative nonparametric approach is to leave the distribution of z_i unspecified. In either case, the distribution of z_i may be approximated by a discrete distribution with finite support. Then the likelihood function $L^*(\boldsymbol{\beta})$ for the GLMwRE and its approximation $L(\boldsymbol{\beta})$ can

This paper was published as a part of the proceedings of the 31st International Workshop on Statistical Modelling, INSA Rennes, 4–8 July 2016. The copyright remains with the author(s). Permission to reproduce or extract any parts of this abstract should be requested from the author(s).

be written as (Aitkin et al., 2009)

$$\begin{aligned} L^*(\boldsymbol{\beta}) &= \prod_{i=1}^{n} \int f(y_i|\boldsymbol{\beta}, \phi, z_i)\varpi(z_i)dz_i \\ &\approx \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k f(y_i|\boldsymbol{\beta}, \phi, \tilde{z}_k) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k f_{ik} = L(\boldsymbol{\beta}), \end{aligned} \tag{1}$$

where $f(\,\cdot\,)$ is the response density, ϕ is the dispersion parameter, $\varpi(\,\cdot\,)$ is the density of the random effect z_i, $\tilde{z}_k$ are mass points and π_k are mass probabilities. From (1) we have an approximate linear predictor for the k-th component of the i-th observation as $g(\mu_{ik}) = \eta_{ik} = \mathbf{x}_i^\top \boldsymbol{\beta} + \tilde{z}_k$ where $\mu_{ik} = \mathrm{E}[y_i|z_i = \tilde{z}_k]$. Let $\ddot{\mathbf{y}}^\top = (\mathbf{y}^\top, \mathbf{y}^\top, \ldots, \mathbf{y}^\top)$ be a vector of nK-dimension of pseudo-observations and the corresponding stacked linear predictor be

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \ddot{\mathbf{X}}\boldsymbol{\beta} + \ddot{\mathbf{z}} \tag{2}$$

where $\boldsymbol{\mu}^\top = (\mu_{11}, \ldots, \mu_{n1}, \ldots, \mu_{1K}, \ldots, \mu_{nK})$, $\boldsymbol{\eta}^\top = (\eta_{11}, \ldots, \eta_{n1}, \ldots, \eta_{1K}, \ldots, \eta_{nK})$, $\ddot{\mathbf{z}}^\top = (\tilde{z}_1, \ldots, \tilde{z}_1, \ldots, \tilde{z}_K, \ldots, \tilde{z}_K)$ is the n times stacked mass point vector, and $\ddot{\mathbf{X}}^\top = (\mathbf{X}^\top, \ldots, \mathbf{X}^\top)$ is the $nK \times p$ pseudo model matrix, where $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$. Maximum Likelihood Estimation (MLE) typically proceeds via the EM algorithm. In the non-parametric approach, π_k and z_k are estimated adaptively along with $\boldsymbol{\beta}$ in the M step and this is known as non-parametric maximum likelihood (NPML). Tabulated Gaussian quadrature points are used for π_k and z_k in the case of Gaussian random effects (the latter being scaled by a parameter σ which needs estimation).

2 The gradient test

The problem considered is that of testing a composite hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^{(0)}$ against a composite alternative $\mathcal{H}_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_1^{(0)}$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, $\boldsymbol{\beta}_1 = (\beta_1, \ldots, \beta_q)^\top$ is a $q-$dimensional parameter of interest with $q \leqslant p$, $\boldsymbol{\beta}_2 = (\beta_{q+1}, \ldots, \beta_p)^\top$ is a $(p-q)-$dimensional nuisance parameter and $\boldsymbol{\beta}_1^{(0)}$ is a specified vector. This induces the partitioning $\ddot{\mathbf{X}} = (\ddot{\mathbf{X}}_1, \ddot{\mathbf{X}}_2)$. Let

$$\mathbf{U}(\boldsymbol{\beta}) = \partial \log L(\boldsymbol{\beta})/\partial \boldsymbol{\beta} = \{\mathbf{U}_1^\top(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2), \mathbf{U}_2^\top(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\}^\top = \{\mathbf{U}_1^\top, \mathbf{U}_2^\top\}^\top$$

be the corresponding partition of the total score function for $\boldsymbol{\beta}$. The unrestricted MLE of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1^\top, \hat{\boldsymbol{\beta}}_2^\top)^\top$ and the restricted MLE of $\boldsymbol{\beta}_2$ under $\mathcal{H}_0$ is written $\tilde{\boldsymbol{\beta}}_2$. Functions evaluated at the point $\tilde{\boldsymbol{\beta}}^\top = (\boldsymbol{\beta}_1^{(0)\top}, \tilde{\boldsymbol{\beta}}_2^\top)$ will be distinguished by the addition of a tilde. The gradient statistic $\xi_{\mathcal{T}}$ for testing $\mathcal{H}_0$ versus $\mathcal{H}_1$ has the simple form $\xi_{\mathcal{T}} = \tilde{\mathbf{U}}_1^\top(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)})$ (Terrell, 2002). In the context and notation set out earlier, one has $\tilde{\mathbf{U}}_1 = \ddot{\mathbf{X}}_1^\top \tilde{\mathbf{D}}(\ddot{\mathbf{y}} - \tilde{\boldsymbol{\mu}})$ and $\mathbf{D}$ is the diagonal matrix with diagonal entries $d_{11}, \ldots, d_{n1}, \ldots, d_{1K}, \ldots, d_{nK}$

given by $d_{ik} = (\phi\omega_{ik}/V_{ik})(d\mu_{ik}/d\eta_{ik})$ where $\omega_{ik} = \pi_k f_{ik}/\sum_{l=1}^{K} \pi_l f_{il}$ and V_{ik} is the variance function applied to μ_{ik}. Therefore, the gradient statistic formula for testing $\mathcal{H}_0$ is

$$\xi_{\mathcal{T}} = (\ddot{\mathbf{y}} - \tilde{\boldsymbol{\mu}})^\top \tilde{\mathbf{D}} \ddot{\mathbf{X}}_1 (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)}). \tag{3}$$

Based on Terrell's (2002) results, the distribution of $\xi_{\mathcal{T}}$ tends under $\mathcal{H}_0$ to the $\chi^2(q)$ distribution as n increases. Theoretically, the $\xi_{\mathcal{T}}$, likelihood-ratio (LR) $\xi_{\mathcal{LR}}$, Wald $\xi_{\mathcal{W}}$ and Rao $\xi_{\mathcal{R}}$ statistics are asymptotically equivalent since they all have the same asymptotic distribution under $\mathcal{H}_0$. Nonetheless, for finite samples the size and/or power of the tests may differ. Consequently, we provide numerical simulation results to compare their performance.

3 Simulation experiment

We report results of Monte Carlo simulations assessing properties of $\xi_{\mathcal{T}}$ in finite samples. For this, we establish a model with linear predictor

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + z_i, \text{ for } i = 1, \ldots, n$$

where x_1, x_3 and x_4 are samples of size n from $\mathcal{U}(0,1)$, $\mathcal{F}(2,5)$ and $t(3)$, respectively. The parameter values are β_0=1, β_1=−1, $\beta_{2i} = (i \mod 3) - 1$ and $\phi = 1$. The random effects z_i are samples from $\mathcal{N}(0, 8^{-2})$ for the Gaussian quadrature fitting and from a discrete distribution which takes K values from $\mathcal{N}(0, 8^{-2})$ and probabilities from $\mathcal{U}(0,1)$ for the NPML fitting. The simulation results are based on Normal with identity link and Poisson and Gamma models with log link function. We took samples of 50, 100, 200 and 400 observations and the number of replications was 10,000 and $K = 3$. Our aim is to test $\mathcal{H}_0 : (\beta_3, \beta_4)^\top = (0,0)^\top$ versus $\mathcal{H}_1 : (\beta_3, \beta_4)^\top \neq (0,0)^\top$. Table 1 shows the null rejection rates of each test for two response distributions. Overall, the gradient statistic has rejection rates closer to the nominal levels. We set $n = 400$, $K = 3$ and $\alpha = 5\%$ for the power simulations where we computed the rejection rates under the alternative hypothesis $\beta_3 = \beta_4 = \delta$, for $-4 \leq \delta \leq 4$. Figure 1 shows that the power curves for $\xi_{\mathcal{LR}}$ and $\xi_{\mathcal{T}}$ are practically identical and that $\xi_{\mathcal{W}}$ and $\xi_{\mathcal{R}}$ have rather unusual curves, especially for the NPML model.

4 Concluding remarks

The gradient test shows itself as a useful inferential tool in the context of GLMwRE for several reasons. Firstly, its statistic requires neither the Fisher information matrix nor its inverse, which is an important simplification compared to the Wald and Rao statistics. Secondly according to our

TABLE 1. Null rejection rates (%).

	n	α	Gaussian quadrature $\xi_{\mathcal{LR}}$	Gaussian quadrature $\xi_{\mathcal{W}}$	Gaussian quadrature $\xi_{\mathcal{R}}$	Gaussian quadrature $\xi_{\mathcal{T}}$	NPML $\xi_{\mathcal{LR}}$	NPML $\xi_{\mathcal{W}}$	NPML $\xi_{\mathcal{R}}$	NPML $\xi_{\mathcal{T}}$
Normal	50	10	13.36	16.10	10.67	11.94	45.97	80.56	3.66	25.00
		5	7.12	9.52	5.24	6.06	33.62	76.57	1.91	16.19
		1	1.78	2.97	0.93	1.12	15.46	68.85	0.48	5.34
	100	10	11.73	12.99	10.51	11.16	24.90	60.19	4.64	17.08
		5	6.08	7.11	5.22	5.59	15.78	53.86	2.58	9.83
		1	1.25	1.72	0.96	1.08	5.18	43.30	0.74	2.50
	200	10	11.45	12.24	10.62	11.16	15.60	38.70	6.53	13.23
		5	5.88	6.49	5.24	5.55	8.58	30.89	3.75	7.09
		1	1.21	1.48	1.02	1.08	2.47	19.79	1.23	1.72
	400	10	10.47	10.95	9.98	10.32	12.78	23.99	9.35	11.96
		5	5.36	5.86	5.04	5.24	6.66	17.20	5.33	6.19
		1	1.15	1.29	0.99	1.09	1.53	8.31	1.70	1.25
Poisson	50	10	10.11	11.92	7.90	10.48	9.36	4.89	16.50	8.91
		5	5.01	6.70	3.86	5.46	4.50	2.24	9.64	4.04
		1	1.13	1.74	0.83	1.40	0.73	0.42	2.82	0.60
	100	10	10.32	12.15	8.56	10.50	9.98	5.31	16.78	9.57
		5	5.20	6.51	4.18	5.43	4.97	2.49	10.06	4.75
		1	1.15	1.65	0.78	1.34	0.88	0.46	3.28	0.92
	200	10	10.45	11.77	8.53	10.72	10.06	5.59	17.77	9.88
		5	4.98	6.20	4.17	5.22	5.05	2.65	10.80	4.88
		1	0.95	1.47	0.74	1.12	0.93	0.52	3.25	1.01
	400	10	9.68	11.15	8.25	9.82	9.50	5.06	16.52	9.68
		5	4.86	5.93	4.23	4.97	4.64	2.13	9.89	4.64
		1	0.97	1.41	0.77	1.04	0.87	0.46	2.83	0.93
Gamma	50	10	13.81	24.13	12.27	15.77	37.47	68.51	7.06	27.24
		5	7.73	16.51	6.79	8.29	27.98	62.20	3.53	17.65
		1	2.07	7.37	1.77	1.89	13.62	51.27	0.93	6.31
	100	10	11.98	18.52	11.62	13.08	22.97	52.04	5.31	18.65
		5	6.39	11.66	6.45	6.49	15.10	43.99	2.89	10.68
		1	1.54	4.49	2.05	1.39	6.06	31.10	0.90	2.84
	200	10	10.78	15.65	10.23	11.24	16.48	38.57	5.14	13.79
		5	5.33	9.33	5.37	5.49	10.45	29.59	2.87	7.59
		1	1.18	2.82	1.64	1.12	3.60	16.98	0.78	1.58
	400	10	10.38	13.21	9.80	10.49	14.10	28.43	5.72	12.22
		5	5.17	7.85	5.17	5.20	8.04	20.40	3.02	6.00
		1	1.08	1.99	1.43	0.91	2.63	10.04	0.77	1.28

simulations, the null rejection rates of the gradient test are much closer to the true nominal levels than the other three tests for the normal response

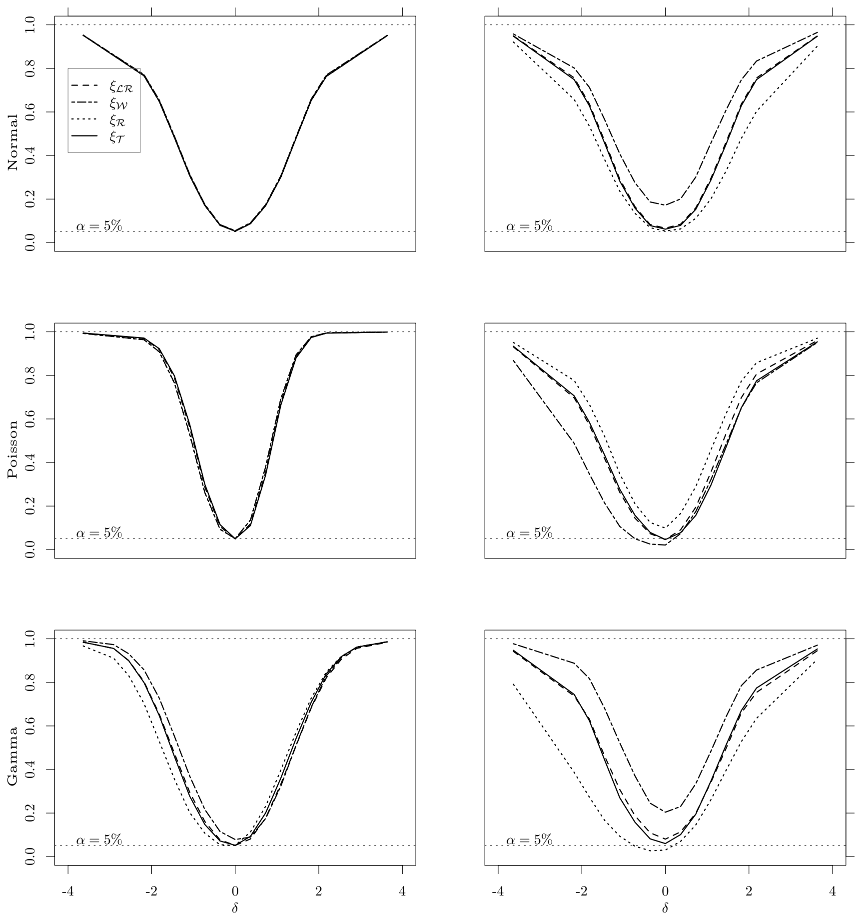

FIGURE 1. Power of the four tests: $n = 400$, $k = 3$, $\alpha = 5\%$. Left, for Gaussian quadrature fitting and right, for NPML fitting.

model and both gradient and LR tests have good rates for the Poisson response. Finally, our power simulations suggest that the gradient and LR tests have similar power properties. In sum, this indicates that the gradient tests should be preferred in the context of GLMwRE.

Acknowledgments: We gratefully acknowledge grant n° 9622/13-6 from CAPES foundation through Brazil's Science Without Borders Program.

References

Aitkin, M.A., Francis, B., Hinde, J. and Darnell, R. (2009). *Statistical Modelling in R.* Oxford: Oxford University Press.

Terrell, G.R. (2002). The gradient statistic. *Computing Science and Statistics*, **34** 206 – 215.